Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Overview of the LES ARM Symbiotic Simulation and Observations (LASSO) Workflow

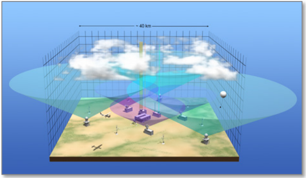
William I. Gustafson Jr.[1] and Andrew M. Vogelmann[2]
and the members of our team

[1]Pacific Northwest National Laboratory, [2]Brookhaven National Laboratory

U.S. DEPARTMENT OF ENERGY

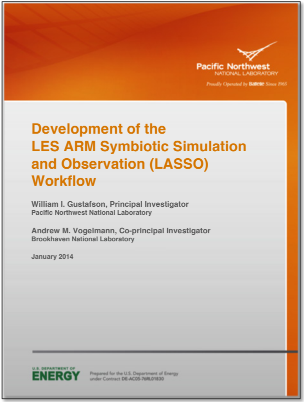
ARM isn't just observations anymore

- ARM has provided abundant observations of climate-relevant variables for over 20 years, resulting in an abundance of new knowledge
- Observations can only get us so far; fully understanding processes often requires a modeling component to supplement observations
  - Not everything can be measured
  - The spatial scale of things that can be measured is not always optimal
- We are starting a 2-year pilot project to develop the infrastructure needed for routine LES modeling by ARM



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle *Since 1965*

The LASSO Project



- **LES ARM Symbiotic Simulation and Observations (LASSO) Workflow**

- Primary team members
  - PNNL: William Gustafson, Larry Berg, Jerome Fast, Mikhail Ovchinnikov, Heng Xiao
  - BNL: Andy Vogelmann, Satoshi Endo, Ed Luke, Tami Toto
  - UCLA: Zhijin Li

- Supporting team members
  - PNNL: Sherman Beus, Jennifer Comstock, Zhe Feng, Rob Newsom, Laura Riihimaki, Chitra Sivaraman
  - BNL: Alice Cialella, Scott Giangrande, Mike Jensen, Karen Johnson, Pavlos Kollias, Wuyin Lin, Yangan Liu
  - NOAA: Dave Turner

The grand vision

LASSO will create a powerful new capability for furthering ARM's mission to advance understanding of cloud, radiation, aerosol, and land-surface processes.

The combined observational and modeling elements will enable a new level of scientific inquiry by connecting processes and context to observations and providing needed statistics for details that cannot be measured.

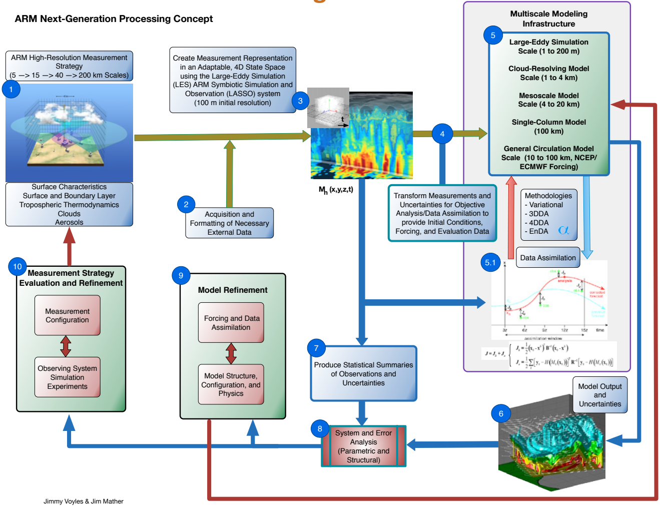
The result will be improved process understanding that facilitates concomitant improvements in climate model parameterizations.

Practical interpretation of the vision

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- **ARM will generate routine simulations for selected events, which will evolve into a database of simulations available to the scientific community**
  - Provides information to robustly evaluate process understanding and parameterizations
  - Avoids the limitation of a small number of well behaved cases
  - Overcomes individual PI's general ability to simulate only a handful of cases
- **Initial focus will be shallow convection at the SGP site**
  - ShCu potentially play a role in the warm temperature bias of GCMs over the Great Plains
  - Addresses increasing community interest in understanding land heterogeneity impact on ShCu and other clouds, along with how to parameterize this effect
  - New PBL/Cu parameterizations need long-term statistics of boundary layer and cloud characteristics for tuning and evaluation

Initial vision for ARM modeling



Jimmy Voyles & Jim Mather
July 2014

Will be developing new products based on large-eddy simulations (LES)

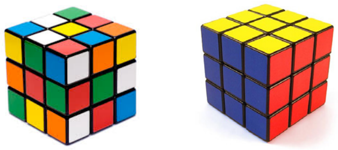
- **Primary product will be multi-dimensional "data cubes"**
  - A merged observation-model approach to describe the most likely conditions at the megasite
  - "Cube" is a misnomer

- **Added value will come from the steps needed to generate the cubes**
  - Categorization of daily weather conditions
  - Ensemble of forcing data
  - Tools to access and analyze ARM data

What is a "data cube"? (1)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

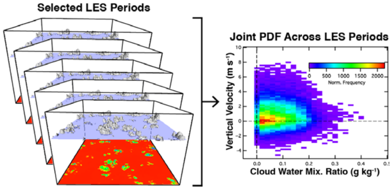
- A statistically guided best representation of the atmospheric conditions above the megasite.

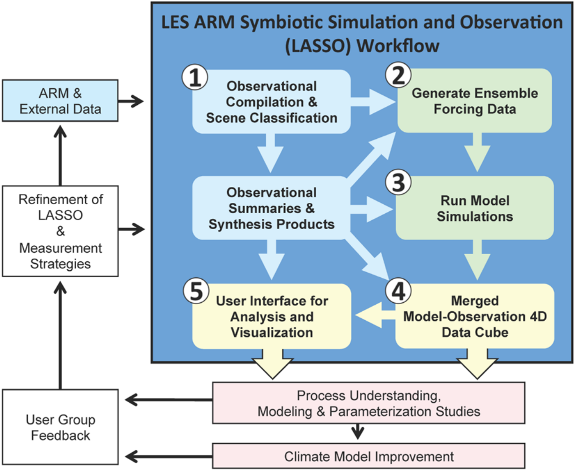


- A combination of model output, observations, and standardized metrics. Models are unable to represent every cloud and the exact heterogeneity within the domain. Likewise, observations are unable to give a complete representation of the volume.

What is a "data cube"? (2)

- The combined model-obs package will be as much of an interface as a physical combination
  - Tools to easily access cube contents both from within the ARM archive and externally
  - Ability to compare multiple data streams within an event as well as across events
  - Ability to use common grids and time sampling, where appropriate




LES ARM Symbiotic Simulation and Observation (LASSO) Workflow
ARM & External Data
Refinement of LASSO & Measurement Strategies
User Group Feedback
1 Observational Compilation & Scene Classification
2 Generate Ensemble Forcing Data
Observational Summaries & Synthesis Products
3 Run Model Simulations
5 User Interface for Analysis and Visualization
4 Merged Model-Observation 4D Data Cube
Process Understanding, Modeling & Parameterization Studies
Climate Model Improvement

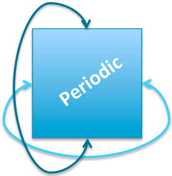
1) Observation compilation and scene selection

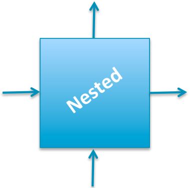
- Initial focus on shallow convection
  - Existing climatologies show 10-20% of spring and summer days have ShCu based on strict definitions of relevant days, e.g. only ShCu present
    (Berg and Kassianov, 2008; Chandra et al., 2013; Zhang & Klein, 2013)
  - Relaxing the criteria will increase the number of selected event days

- Workflow needs
  - An automated cloud and weather regime identification product
  - Generation of searchable metrics

2) Generate ensemble forcing data

- Type of forcing depends on the domain configuration





- Requires a column-based forcing applied uniformly throughout the domain
- Requires a spatially varying forcing applied along the boundaries

2) Generate ensemble forcing data

- Use a forcing ensemble to assist with estimating forcing uncertainty

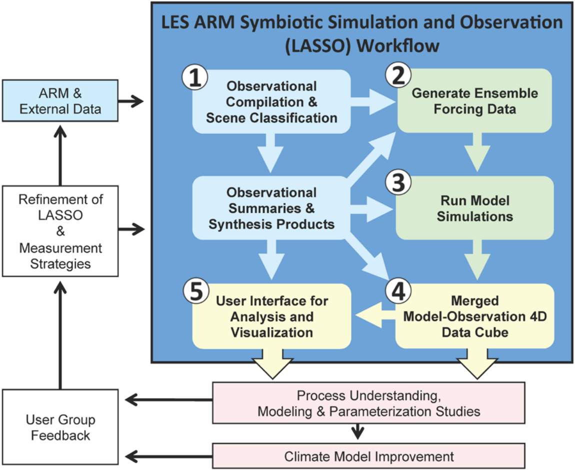
- Have proposed using 3 forcing methodologies for initial testing
  - ARM continuous forcing data based on constrained variational-analysis approach (Xie et al., 2004)
  - ECMWF forecast based forcings
  - Multi-scale Data Assimilation (MS-DA) methodology using 3DVar DA, with additional tests using ensemble Kalman filter (EnKF) DA (Li et al, 2013)
    - Multiple versions based on GFS, NAM, HRRR, NARR
    - EnKF could also provide ensemble scenarios
  - Open to testing additional forcing methodologies as they become available and time permits

2) Generate ensemble forcing data

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Forcings can potentially be generated and stored for all days, not just simulated days

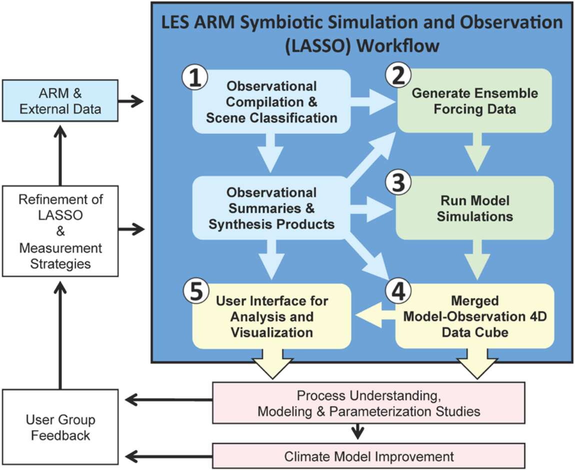
- Workflow needs
  - Ingestion of a wide range of weather forecast datasets
  - ARM data/VAPs available within days to weeks of real time
  - Automated generation of each forcing dataset
    - Doing our own data assimilation essentially entails running our own weather forecast system based on data assimilation software from Zhijin Li
    - Some methodologies have been hampered by the need for user input


LES ARM Symbiotic Simulation and Observation (LASSO) Workflow
ARM & External Data
Refinement of LASSO & Measurement Strategies
User Group Feedback
1
Observational Compilation & Scene Classification
2
Generate Ensemble Forcing Data
Observational Summaries & Synthesis Products
3
Run Model Simulations
5
User Interface for Analysis and Visualization
4
Merged Model-Observation 4D Data Cube
Process Understanding, Modeling & Parameterization Studies
Climate Model Improvement

3) Run model simulations

- Will test two candidate models (WRF and SAM) for periodic and nested boundary configurations, as appropriate
- Ensemble of sims with bulk microphysics based on multiple forcings
- One simulation with spectral bin microphysics
- Users can conduct their own additional sensitivity simulations using the forcing conditions and ARM data

- Workflow needs
  - Automated workflow for running the model
  - Substantial computing resources
    - O(1000-10,000) compute cores
    - O(1) PB storage per year
  - Fully scriptable tools to remotely retrieve ARM data in realistic timeframes


LES ARM Symbiotic Simulation and Observation (LASSO) Workflow
1
Observational Compilation & Scene Classification
2
Generate Ensemble Forcing Data
Observational Summaries & Synthesis Products
3
Run Model Simulations
5
User Interface for Analysis and Visualization
4
Merged Model-Observation 4D Data Cube
ARM & External Data
Refinement of LASSO & Measurement Strategies
User Group Feedback
Process Understanding, Modeling & Parameterization Studies
Climate Model Improvement

4) Merged model-observation data cube

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- This step generates
  - Metadata useful for documenting the event and later searches
  - Metrics and statistics for the ensemble and model-observation comparisons
  - Subsetting of model output, and post-processing steps for data that will be stored in the ARM Archive (e.g., generating model output comparable to radar observations)
- Some data post-processing will be "on-the-fly" via tools like ADI that can do temporal and spatial sampling. More expensive and technically complicated post-processing will need to be pre-calculated before archiving.
- Our project is tasked with defining the needs and will work with ARM developers to determine how to best meet the needs.

4) Merged model-observation 4-D data cube

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

- Example metadata & metrics
  - Information for finding events such as weather characteristics, min/max/mean/std. dev. of selected variables, availability of key instrumentation
  - Taylor diagrams to diagnose simulation accuracy
  - Spaghetti diagrams to evaluate ensemble spread and outliers

- Workflow needs
  - Integration with ADI and possible enhancements to accommodate specific needs
  - Efficient data access and software to generate the data cube
  - Implementation of the post-processing into ARM's overall workflow

5) User interface for analysis & visualization

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

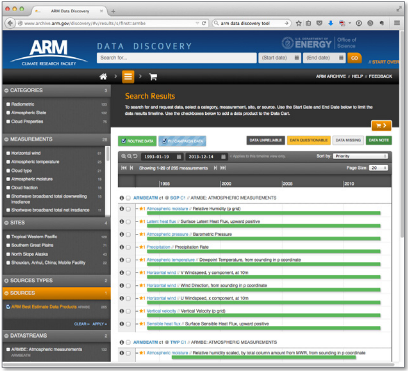
- Our goal is to streamline observation-model intercomparison
- The dataflow will be overwhelming and will require powerful tools that are both technically smart and intuitive to use
- Our project is tasked with defining priorities and possible methodologies to be considered for implementation by ARM programmers
  - Web interface for data discovery
  - Data cube analysis from within the ARM cluster
  - Analysis tools for external access to the archive

Photo by Walter Parenteau - http://flic.kr/p/EbjnM

What we are trying to avoid…

Data cube web interface

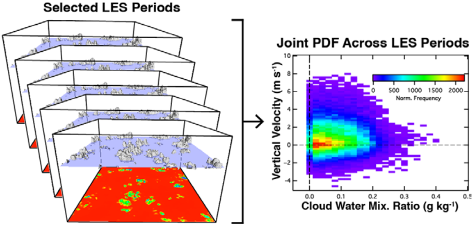
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*



- Designed for event selection and simple analyses
- Search capability tied to pre-calculated metadata and metrics
- Displays quick-look plots to easily scan through events
- User selected comparisons between model and observations
- Ability to request subsetted, packaged data

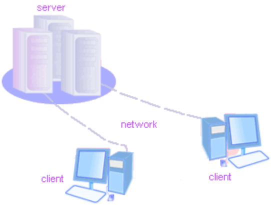
Data cube analysis tools from within ARM cluster

- Free-form interactive capabilities for analysis and visualization
  - Ability to use Python, Fortran, IDL, Matlab, etc. directly on ARM observations and simulations
  - Evolving library of analysis tools, e.g., Py-ART, instrument simulators
- Ability to run SCMs using the same forcing as the LES and then do comparisons between the two, similar to the KNMI Parameterization Testbed



Data cube tools for external access to archive

- Software tools to remotely query and retrieve data cube contents
- Take advantage of server-client architecture where appropriate
  - Use ADI to regrid and sub-sample data at the archive before sending to the user's computer
  - Pre-conglomeration of variables into bundles to facilitate later comparison
- Ability to stage data necessary to perform simulations complimentary to ARM's simulations
  - Enable users to use the same workflow as ARM, but with their own model configuration
- Tied to best-in-practice technologies for data transfer, e.g., Globus file transfers



5) User interface for analysis & visualization

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Workflow needs
  - Web interface to easily discover and work with data cube information
  - Enhancements to the Data Discovery Tool to include additional searches on event metadata
  - Software libraries to aid automated data access and analysis (server-client techniques would be great)
  - A way for users to reproduce ARM's LASSO workflow at their home institution
  - Open source software for community users
  - Tools to facilitate easy, scriptable access to the ARM archive via Globus
  - Optimization of data storage for quick access
  - ARM computing resources made available to users

Technical issues to be addressed

- What model should we use?
- What is the optimal domain configuration to meet the most scientific objectives?
  - Horizontal and vertical extent, resolution, boundary type (nested or periodic)
- How far can a traditional (periodic) domain be pushed, e.g., non-ShCu events?
- How consistently will a nested LES approach work?
- How does one identify good vs. bad forcing datasets?
- What is the optimal modeling infrastructure that can be ported to additional sites and weather regimes?
- How does one effectively initialize and use an interactive soil model at LES scales when observations are sparse?
- How does one initialize the LES when clouds are present?
- Is there enough value in bin microphysics to justify the added cost?
- How many ensemble members is appropriate at LES scales?
- How does one keep all the constituents happy and engaged?

(Technically, the last question isn't technical, but it is critical.)

User feedback is critical for ARM's success

- What science do you want to do with the ARM data cubes that would impact the event selection criteria?
- How would you use ARM-generated forcing data?
- Would you use forcing data generated for days other than those simulated by ARM?
- What forcing types do you find most valuable?
- What model configuration options would best serve as a universal control simulation?
- What spatial scales are needed, both horizontally and vertically?
- What variables, statistics, and metrics would be most valuable to you?
- How much of the raw model output would you use?
- How would you want to access ARM's data cubes?
- What ARM-provided analysis and visualization tools would you find useful?

Summary of key deliverables

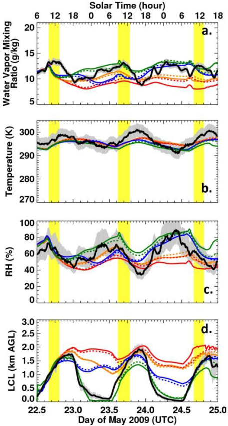
- A prototype workflow with all the components and requirements defined, which will be automated by ARM software developers in coordination with the project principal investigators (PIs). Our team will develop and test the model configuration, domain choice, forcing datasets, and operating scenarios.
- Sets of simulations consisting of multiple model configurations that will be suitable for implementing observation-model integration software; these will span shallow cloud days over multiple seasons.
- Extensive recommendations surrounding the many facets of developing a useful workflow including
  - Implementation of high-resolution modeling constrained by ARM observations
  - Construction of the data cubes to combine observations, model output, and metrics
  - Development of analysis and visualization tools including instrument simulators
  - Efficient user access to LASSO data cubes via ARM infrastructure and externally

The End

Different forcings make a difference

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by **Battelle** *Since 1965*

RACORO case simulated by DHARMA (ARM & ECMWF) and WRF (for MS-DA)



Black = Observations

DHARMA simulations:
Orange = ARM forcing for 300 km scale
Red = ARM forcing for 150 m scale
Blue = ECMWF forcing

WRM simulation:
Green = MS-DA forcing

Solid = No thermodynamic relaxation
Dashed = With 12-h thermodynamic relaxation

(Vogelmann, 2015)